DOCUMENT RESUME

ED 064 319 TM 001 484

AUTHOR Pena, Deagelia M.
TITLE The Assumption of a Markoff Chain Model for Interaction Analysis.
PUB DATE Apr 72
NOTE 42p.; Paper presented at the annual meeting of the American Educational Research Association (Chicago, Illinois, April 1972)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Data Analysis; *Hypothesis Testing; *Interaction Process Analysis; *Mathematical Models; *Probability Theory; Tables (Data)
IDENTIFIERS Darwin (J H); Likelihood Ratio Criterion; *Markoff Chain Model

ABSTRACT

One approach to the problem of analyzing classroom interaction by use of a Markoff chain is presente . The two sections of the paper present (1) a discusion of Hoel's test of order of a Markoff chain, and a Likelihood Ratio Criterion (LRC) for a two-dependent Markoff chain is given, and (2) a discussion of possible adjustments of Darwin's LRC with which to analyze data on a one-dependent chain assumption. The results of Hoel's test of order of a Markoff chain showed that a two-dependent (Order two) model is a better fit to interaction data than the one-dependent model (Order one). Tables and graphs present the statistical data. (DB)

THE ASSUMPTION OF A MARKOFF CHAIN MODEL FOR INTERACTION ANALYSIS

by
Deagelia M. Peña
Detroit Board of Education
Detroit, Michigan 48202

A paper presented to the Advance
Copy Session, Division D, of the annual meeting of the
American Educational Research Association
April 2-7, 1972
Chicago, Illinois

INTRODUCTION

When a set of interaction data based on Flanders' Interaction Category System, FLICS[1] is recorded in a matrix, the purpose is to look at paired sequences of observation (recorded in the "cells" of the matrix). The ten categories are given in Table 1.

Flanders [1] discussed the interpretation of various sections of the matrix which are combinations of the paired sequences.

The concept of an interaction analysis _matrix_ is closely related to a _one-dependent Markoff chain_ (also simply called a Markoff chain). Many researchers are analyzing differences between two or more interaction matrices with the use of a criterion based on a Markoff chain model.

Feller's [2, p. 340] definition will be paraphrased here and subsequently applied to interaction analysis data.

A _Markoff chain_ is a sequence of trials with possible outcomes $E_1, E_2, \ldots,$ in which the probabilities of sample sequences are defined by

$$p\{E_{j_0}, E_{j_1}, \ldots, E_{j_n}\} = a_{j_0} p_{j_0 j_1} p_{j_1 j_2} \cdots p_{j_{n-2} j_{n-1}} p_{j_{n-1} j_n},$$

---

[1] FLICS is an acronym adopted by the author, to refer specifically to Flanders' ten-category system of recording verbal behavior, in a sequence, as they occur in a classroom--the coding taking place at three-second intervals; the data are used for interaction analysis.

TABLE 1

CATEGORIES FOR INTERACTION ANALYSIS

TEACHER TALK	INDIRECT INFLUENCE	1.	Accepts Feeling: accepts and clarifies the feeling tone of the students in a nonthreatening manner. Feelings may be positive or negative. Predicting or recalling feelings are included.
		2.	Praises or Encourages: praises or encourages student action or behavior. Jokes that release tension, not at the expense of another individual, nodding head or saying, "um hm?" or "go on" are included.
		3.	Accepts or Uses Ideas of Student: clarifying, building, or developing ideas suggested by a student. As a teacher brings more of his own ideas into play, shift to category five.
		4.	Asks Questions: asking a question about content or procedure with the intent that a student answer.
	DIRECT INFLUENCE	5.	Lecturing: giving facts or opinions about content or procedure; expressing his own ideas, asking rhetorical questions.
		6.	Giving Directions: directions, commands, or orders to which a student is expected to comply.
		7.	Criticizing or Justifying Authority: statements intended to change student behavior from non-acceptable to acceptable pattern; bawling someone out; stating why the teacher is doing what he is doing; extreme self-reference.
STUDENT TALK		8.	Student Talk--Response: a student makes a predictable response to teacher. Teacher initiates the contact or solicits student statement and sets limits to what the student says.
		9.	Student Talk--Initiation: talk by students which they initiate. Unpredictable statements in response to teacher. Shift from 8 to 9 as student introduces own ideas.
		10.	Silence or Confusion: pauses, short periods of silence and periods of confusion in which communication cannot be understood by the observer.

in terms of an initial probability distribution a_k for the states E_k at time 0 and fixed conditional probabilities p_{jk} of E_k, given that E_j has occurred at the preceding trial.

In interaction analysis this definition means essentially that anywhere in the interaction sequence, the probability that category (state) j will occur depends only on the preceding category i, and not on any other previous categories in the sequence. As an example, the probability that category nine (student talk--initiation) occurs in a classroom depends only on what occurred immediately before the nine. Looking at Table 1, one may guess that a nine is more likely (has a greater probability) to occur, if the preceding category is a three (accepting ideas of student) than if the preceding is a five.

One approach to the problem of analyzing classroom interaction begins by assuming that the sequence of observations is a one-dependent Markoff Chain. The categories are the "outcomes," $E_1, E_2, \ldots, E_{10}$ in the definition. The initial probability distribution a_k is the probability of initial occurrence of any of the 10 categories. In a classroom, however, it is logical to assume that silence (or confusion) is always the initial state; this makes a_k in the definition equal to <u>one</u> for interaction analysis data.

The conditional probability p_{jk} is the probability of occurrence of category k, given that j is the preceding category.

Many researchers are analyzing differences between two or more matrices for statistical significance by applying Darwin's [3, p. 413]. Assuming a Markoff chain for interaction data, Darwin considered testing the hypothesis that t sets[2] of values p_{ij} (p_{ij} unknown, and $i,j = 1,\ldots,s$, where s equals the number of categories) are equal; that is, two or more matrices have the same p_{ij} for a given i and j. The data to which this refers are t matrices with long sequences. The Likelihood Ratio criterion to test the hypothesis is [3, p. 413]

$$2[\Sigma n_{ijh} \log n_{ijh} - \Sigma n_{i.h} \log n_{i.h} - \Sigma n_{ij.} + \Sigma n_{i..} \log n_{i..}] \quad (1)$$

distributed as chi square with $i,j = 1,\ldots s$, $h = 1,\ldots,t$, and $s(t-1)(s-1)$ degrees of freedom. This criterion when applied to interaction data, was giving results which were too significant; that is, it was too sensitive to slight differences between sets of inter-

---

[2] For those who refer to Darwin [3], the notation t is used instead of r, to avoid confusion with the use of r which denotes the order of a Markoff chain.

action data. The objectives in this inquiry were:

1. to test the order of dependence of the interaction chain and,

2. on the assumption of a one-dependent Markoff chain,
   a. to estimate empirically the power of Darwin's criterion as applied to two composite sequences, and
   b. to arrive at an application of Darwin's criterion, which will reflect educational significance.

This paper has two sections. In Section I Hoel's [4] test of order of a Markoff chain will be discussed. The test showed that a two-dependent (Order two) model is a better fit to interaction data than the one-dependent model (Order one)--a model that is assumed when researchers use Darwin's criterion. Hence, a Likelihood Ratio Criterion (LRC) for a two-dependent Markoff chain will also be presented.

In Section II the author will discuss possible adjustments of Darwin's LRC, if the researcher wishes to analyze data on a one-dependent chain assumption.

In order to do this, a specific alternative hypothesis that two interaction sequences are not equal will be used to calculate the power of Darwin's criterion, given such inequality. That is, it will be shown that, given two matrices which are not equal only by some chance and not due to differences in classroom interaction, Darwin's LRC will "reject" the null hypothesis of equality 100% of the time, when applied to 500 pairs of matrices generated from the two matrices.

From the generated pairs of matrices an empirical distribution of Darwin's LRC will be derived in order to determine a cut-off point for rejection of the null hypothesis, such that rejection would have not only statistical significance but also educational significance.

Sections I and II may be read independently of each other.

Section I

Hoel's Test of Order of Dependence

The data used in this section came from the data bank of Flanders [5]. These were the sixth grade interaction sequences of 30 classrooms on five different subject areas or activities.

Hypotheses

In general notation, the null hypotheses being tested are expressed as follows:

<u>The transition probability</u>, $p_{ij\ldots kl}$, <u>in an r-step chain is equal to the transition probability</u>, $p_{.j\ldots kl}$, <u>in an (r - 1)-step chain.</u> Thus:

$$H_0: \; p_{ij\ldots kl} = p_{.j\ldots kl}, \qquad i = 1,2,\ldots,s.$$

In this paper we will let $\hat{p}_{ij\ldots kl} = n_{ij\ldots kl}/n_{ij\ldots k.}$ be the Maximum Likelihood estimator of $p_{ij\ldots kl}$. If the hypothesis is true, then the Likelihood Ratio criterion, $-2 \log \lambda$ given below,[3] is, for large samples, distributed as a χ^2 with $s^{r-1}(s - 1)^2$ degrees of freedom:

$$\chi^2 = -2 \log \lambda = 2 \sum_{i,\ldots,l} n_{ij\ldots kl} \left[\log \frac{n_{ij\ldots kl}}{n_{ij\ldots k.}} - \log \frac{n_{.j\ldots kl}}{n_{.j\ldots k.}} \right] \tag{2}$$

In particular, the two null hypotheses tested together with their appropriate χ^2's were:

(1) $H_0: \; p_{ijk} = p_{.jk}, \qquad i = 1,\ldots,10$

---

[3]The logarithms in the formulas throughout this paper are all to the base e.

where

$$\hat{p}_{ijk} = \frac{n_{ijk}}{n_{ij.}}$$

$$\chi^2_{(2)} = 2 \sum_{ijk} n_{ijk} \left(\log \frac{n_{ijk}}{n_{ij.}} - \log \frac{n_{.jk}}{n_{.j.}}\right) \tag{3}$$

(2) H_0: $p_{ij} = p_{.j}$

where

$$\hat{p}_{ij} = \frac{n_{ij}}{n_{i.}}$$

$$\chi^2_{(1)} = 2 \sum_{ij} n_{ij} \left(\log \frac{n_{ij}}{n_{i.}} - \log \frac{n_{.j}}{n_{..}}\right) \quad . \tag{4}$$

The χ^2 subscript refers to the order of dependence r, being tested against an (r - 1)-dependence, and, from what follows, to say that $\chi^2_{(r)}$ is not significant means that an (r - 1)-dependence is as good as an r-dependence assumption. Therefore, an (r - 1)-dependence would be preferable.

The basic idea behind Hoel's test is that a sufficiently large r is chosen and testing is done successively, decreasing r by one each time until a point is reached when r is not significant and r - 1 is significant. This is so because, if a chain is Markovian and the length of the dependence is r + 1 (i.e., the chain is r step and the dependence extends over r + 1 consecutive variables),

$\chi^2_{(r)}$, $\chi^2_{(r+1)}$ . . . will not be significant in increasing α levels, while $\chi^2_{(r-1)}$, $\chi^2_{(r-2)}$ $\cdots$ $\chi^2_{(1)}$ will be significant, in decreasing α levels.

The data did not reasonably allow a test beyond $r = 2$. The test was made, not so much to locate the order of dependence as to have a basis of choice between an $r = 2$ and an $r = 1$ assumption. The result of Hoel's test would be conclusive as to <u>best</u> fit only if $\chi^2_{(r)}$ would not turn out to be significant while $\chi^2_{(r-1)}$ was significant. In this case it would mean that an (r-1) dependence assumption would suffice, while an (r-2) dependence wouldn't. However, if both r and (r-1) dependence were significant, the judgment as to whether an r-dependent model was a better fit than an (r-1) dependence would be discretionary, after the magnitude of difference in the α levels was observed.

Application of the Test

Hoel's test was run on each of the matrices of the 30 teachers, as well as on each of the five activities. The matrices of the activities were formed as follows: Five teachers were chosen at random and their interaction data separated into five sequences (five matrices), one for each of five activities in which they were observed. All five matrices of each activity were combined, yielding one matrix for each activity. The five activities were administrative routine, language arts, social studies, mathematics, and science.

Results of Hoel's Test

Table 2 summarizes the results of the χ^2 tests on each teacher's matrix, and Table 3 those on each of the five activities. The z-column in Table 2 shows that all the 30 χ^2's were significant with $\alpha < .001$ for twenty nine teachers and $\alpha < .01$ for one teacher (Teacher 19). The z-column in Table 3 shows that for $r = 2$ the administrative routine and science chains (across the same five teachers in the other activity chains) were not significant. The results on these two activities (administrative routine and science) may not be true reflections of the actual length of dependence because the table shows that the two chains were relatively much shorter than those of the three other activities.

The z values ($z = \sqrt{2\chi^2} - \sqrt{2df-1}$; see footnote of Table 2) for $r = 2$ for the 30 teachers range from 3.34 to 22.67, while those for $r = 1$ range from 77.4 to 130.23. Hence, a one-dependent chain was still a bad fit and zero dependence should definitely not be considered at all. The marked decrease in z from the assumption of one-dependence to the assumption of two-dependence, may indicate that had it been possible to test the chain for $r = 3$

TABLE 2

TEST OF ORDER OF MARKOFF CHAIN ON 30 TEACHERS

Teacher No.	Length of Chain (No. of Tallies)	r	χ^2	df	z^a
00	6019	2	1415.11	810	12.96
		1	5986.71	81	96.73
08	8388	2	1618.01	810	16.65
		1	10213.53	81	130.23
12	6265	2	1516.75	810	14.84
		1	6085.61	81	97.63
13	5781	2	1306.63	810	10.88
		1	7162.28	81	107.00
15	7080	2	1476.84	810	14.11
		1	7618.80	81	110.75
19	6858	2	945.10	810	3.24
		1	8413.77	81	117.03
24	8117	2	1826.22	810	20.20
		1	8042.57	81	114.14
26	6323	2	1339.10	810	11.51
		1	5814.53	81	95.15
27	6557	2	1613.09	810	16.56
		1	5733.54	81	94.40
28	8360	2	1718.08	810	18.38
		1	9398.20	81	124.41
30	7886	2	1239.17	810	9.55
		1	7478.43	81	109.61
34	6339	2	1972.84	810	22.58
		1	4639.90	81	83.64
37	6386	2	1567.70	810	15.76
		1	4070.33	81	77.54
40	7216	2	1565.61	810	15.72
		1	6492.02	81	101.26
42	6175	2	1538.02	810	15.23
		1	5854.98	81	95.52

TABLE 2-- Continued

Teacher No.	Length of Chain (No. of Tallies)	r	χ^2	df	z^a
48	7103	2	1298.55	810	10.72
		1	6839.02	81	104.26
50	7233	2	1403.73	810	12.75
		1	5958.53	81	96.48
51	8884	2	1593.88	810	16.22
		1	9294.03	81	123.65
53	7462	2	1129.70	810	7.30
		1	5140.46	81	88.70
54	7903	2	1734.55	810	18.66
		1	6315.20	81	99.70
64	6544	2	1279.30	810	10.35
		1	6478.09	81	101.14
72	6831	2	1748.34	810	18.90
		1	6725.48	81	103.30
73	6086	2	1313.20	810	11.01
		1	4216.91	81	79.15
75	5586	2	1104.47	810	6.76
		1	6557.90	81	101.84
77	7201	2	1376.54	810	12.23
		1	5510.03	81	92.29
80	9075	2	1978.74	810	22.67
		1	7030.50	81	105.89
84	6639	2	1454.74	810	13.70
		1	7932.27	81	113.27
89	6789	2	1483.50	810	14.23
		1	7829.44	81	112.45
91	5749	2	1391.83	810	12.52
		1	5068.90	81	88.00
95	6339	2	1698.90	810	18.05
		1	4870.74	81	86.01

[a]Because d.f. is large (> 70), the expression $z = \sqrt{2\chi^2} - \sqrt{2d.f. - 1}$ was used as a normal deviate with unit variance, whereby the probability for χ^2 corresponds with that of a single tail of the normal curve.

against $r = 2$, the z values may not have turned out to be significant, indicating that the best fit on the Markoff chain assumption would be the two-dependent chain model.

TABLE 3

TEST OF ORDER OF MARKOFF CHAIN ON FIVE ACTIVITIES

Chain[a] Description	Length of Chain	r	χ^2	df	z
Adm. routine	2398	2	703.17	810	-2.74[b]
		1	2606.54	81	59.51
Lang. arts	11756	2	3009.29	810	37.34
		1	12507.93	81	145.48
Soc. Studies	7262	2	1882.40	810	21.12
		1	7233.41	81	107.59
Mathematics	8547	2	1487.62	810	14.31
		1	7348.40	81	108.54
Science	2506	2	686.54	810	-3.18[b]
		1	2971.44	81	64.40

[a]The chain is across five teachers (same activity) drawn at random from the 30 teachers in the study.

[b]Not significant.

A Likelihood Ratio Statistic for the Two-Step Model

This part of Section I is mainly expository. It presents a simple application of the Likelihood Ratio test for a Markoff chain of order two. Such a test appears in the literature on the subject. For example, Anderson and Goodman [8, p. 103] obtained a χ^2 test of goodness of fit, i.e., a test of the hypothesis that two samples are from the same Markoff chain of a given order.

Maximum Likelihood (ML) Estimator of p_{ijkh}

Under the two-dependent chain model, the ML estimator of p_{ijk} is [4, p.430]:

$$\hat{p}_{ijk} = \frac{n_{ijk}}{n_{ij.}} \qquad (5)$$

If there are t matrices to be compared, (5) is the ML estimator of the p_{ijk} of each of the t matrices. Hence, for the h^{th} matrix, the ML estimator is

$$\hat{p}_{ijkh} = \frac{n_{ijkh}}{n_{ij.h}} \; . \qquad (6)$$

An Estimator of p_{ijkh} Under the Hypothesis That t Matrices Are Equal

The test statistic to be developed is a test of the hypothesis (H_0) that two or more matrices of interaction data under the two-dependent Markoff chain model are equal.

Let p_{ijkh} be the probability derived from some known distribution of the n_{ijkh} representing the probability for the h^{th} teacher to have an observed interaction fall on the k^{th} category (state), given that two preceding states are i,j. Then

$$H_0\text{:} \; p_{ijk1} = p_{ijk2} = \cdots = p_{ijkt}. \qquad (7)$$

Under the assumption that the interaction data for each teacher are generated by the same two-step Markoff process, the hypothesis in (7) states that the transition probability is the same for each teacher.
Hence, the estimator of p_{ijkh} under H_0 can be equated

with the Maximum Likelihood estimator of p_{ijk} of a two-step Markoff chain, that is,

$$\hat{p}_{ijkh} \ (H_0) = \hat{p}_{ijk} \ . \qquad (8)$$

The hypothesis H_0 in (7) states that the t sets of p_{ijkh} are the same for every combination i, j. There are, therefore, s^2 tables (s x t) of the contingency type. Under this hypothesis,

$$\hat{p}_{ijkh}(H_0) = n_{ijk.}/n_{ij..} \ . \qquad (9)$$

Equation (9) is an estimator for p_{ijk} for any of the t matrices under the hypothesis set up in (7). The ratio of the Likelihood Function (LF) evaluated for the estimator in (9) to that evaluated for the ML estimator of p_{ijkh} in (6) is the Likelihood Ratio. If H_0 of (7) is true, this Likelihood Ratio will not be significantly different from <u>one</u>.

Development of the Likelihood Ratio Statistic

The LF of p_{ijkh} is

$$L = \prod_{ijkh} (p_{ijkh})^{n_{ijkh}} \qquad (10)$$

Let $\hat{p}_{ijkh}$ be the ML estimator of p_{ijkh}, and $\hat{p}_{ijkh(0)}$ be the estimator under H_0. We have the following:

$$\hat{p}_{ijkh} = \frac{n_{ijkh}}{n_{ij.h}} \quad \text{(for each h).} \tag{11}$$

$$\hat{p}_{ijkh(0)} = \hat{p}_{ijk} = \frac{n_{ijk.}}{n_{ij..}} \quad \text{(for any h).} \tag{12}$$

$$\lambda = \frac{L_0(\max)}{L(\max)} = \frac{L\left(\hat{p}_{ijkh(0)}\right)}{L(\hat{p}_{ijkh})} \quad . \tag{13}$$

$$L\left(\hat{p}_{ijkh(0)}\right) = \prod_{ijkh} \left(\frac{n_{ijk.}}{n_{ij..}}\right)^{n_{ijkh}} \quad . \tag{14}$$

$$L(\hat{p}_{ijkh}) = \prod_{ijkh} \left(\frac{n_{ijkh}}{n_{ij.h}}\right)^{n_{ijkh}} \quad . \tag{15}$$

$$\log \lambda = \sum_{ijkh} n_{ijkh}\left(\log \frac{n_{ijk.}}{n_{ij..}} - \log \frac{n_{ijkh}}{n_{ij.h}}\right) \quad . \tag{16}$$

$$-2 \log \lambda = 2\Big(\sum_{ijkh} n_{ijkh} \log n_{ijkh} - \sum_{ijkh} n_{ijkh} \log n_{ij.h}$$

$$- \sum_{ijkh} n_{ijkh} \log n_{ijk.} + \Sigma n_{ijkh} \log n_{ij..}\Big) \quad . \tag{17}$$

The last three summations of equation (17) are

$$\sum_{ijkh} n_{ijkh} \log n_{ij.h} = \sum_{ijh} n_{ij.h} \log n_{ij.h} \tag{18}$$

$$\sum_{ijkh} n_{ijkh} \log n_{ijk.} = \sum_{ijk} n_{ijk.} \log n_{ijk.} \tag{19}$$

and

$$\sum_{ijkh} n_{ijkh} \log n_{ij..} = \sum_{ij} n_{ij..} \log n_{ij..} \qquad (20)$$

the dot taking the place of the summation sign.

Substituting (18), (19), and (20) in (17), we have

$$-2 \log \lambda = 2\Big(\sum_{ijkh} n_{ijkh} \log n_{ijkh} - \sum_{ijh} n_{ij.h} \log n_{ij.h} - \sum_{ijk} n_{ijk.} \log n_{ijk.} + \sum_{ij} n_{ij..} \log n_{ij..}\Big) . \qquad (21)$$

Eq. (21) is the LR test criterion of difference among t matrices, on the assumption of a Markoff chain of order two. Each matrix is identified by the h subscript, $h = 1,\ldots,t$, and n_{ijkh} denotes the frequency of the sequence $X^{(n-2\tau)} = i$, $X^{(n-\tau)} = j$, $X^{(n)} = k$, for teacher h. The dot means that summation has been carried out over the replaced subscript, so that

$$n_{ij.h} = \sum_{k} n_{ijkh}, \quad n_{ijk.} = \sum_{h} n_{ijkh}, \quad \text{and } n_{ij..} = \sum_{kh} n_{ijkh}.$$

It has been shown in literature that $-2 \log \lambda$ (3.37) is distributed asymptotically as χ^2, on $s^2(s-1)(t-1)$ degrees of freedom.

Section II

Analysis of One-Dependent (Interaction) Chains

The Problem of Comparing Two Interaction Sequences

When interaction data are displayed in a 10 x 10 matrix form, the underlying assumption is that these data were generated by a Markoff chain of order one. In order to test the equality of two interaction analysis sequences, one computes Darwin's LRC (Eq. (1)),

$$2[\Sigma n_{jkl} \log n_{jkl} - \Sigma n_{j.l} \log n_{j.l} \quad \Sigma n_{jk.} \log n_{jk.} + \Sigma n_{j..} \log n_{j..}] ,$$

which has a Chi Square distribution for a large n. Then using the Chi Square table, he determines the statistical significance of the difference between the two sets of data.

The results of past studies, however, have shown that the test is so sensitive that small differences between two interaction sequences yield a significant Chi Square. Hence, to interpret the results of the test from an educational viewpoint--that is, to see whether the statistical significance has any practical meaning in education--one needs to find out how often the test rejects the null hypothesis of equality when it is assumed that N pairs of sequences to which Darwin's LRC is applied are generated by a pair of Markoff chain models with known transition probabilities and when it is assumed that the transition probabilities indicate that the sequences in the pair are educationally homogeneous.

The proportion of the times the test would reject the null hypothesis of equality would be an approximate measure of the chance of being right, in the statistical sense, in inferring that the two interaction sequences are different from each other. Since Darwin's test is very powerful, it is possible that all Chi Square tests would reject H_0, i.e., the power of the test equals one. Therefore, one would quite frequently make a statistical inference that the two sets of data, of the type specified, are significantly different. However, since the two sample sequences were generated from two educationally homogeneous matrices, most of the time our inference is wrong in the educational sense. Hence, the interpretation of the statistical significance of Darwin's LRC can be very misleading, especially if one has to make an inference from such a test that one type of teacher behavior is more effective in producing certain educational outcomes than is another. One way to estimate the power of the test would be to study an empirical distribution of N Chi Squares and see how each Chi Square value is associated with an estimated probability of occurrence. In this way, one may evaluate a computed Chi Square in terms of the particular distribution.

In this section, (1) an estimate of the power of Darwin's criterion will be obtained and (2) a cumulative probability distribution of the Chi Square values from pairs of sequences generated from two educationally homogeneous sequences through computer simulation.

The LRC and the Length of Sequence

When two interaction sequences are compared, an estimate of the power of Darwin's test is not the only problem to arise. A secondary problem is an approximation of what happens to the LRC as the length of the sequence (number of tallies) increases. It can be observed that the size of LRC increases with an increase in the number of tallies. The increase in the size of the LRC, in this example, does not relate to the degrees of freedom which, in turn, do not depend on the length of the chain (or number of tallies) but on the number of categories ($s = 10$) and the number of sequences being compared--in this problem, two sequences. The degrees of freedom remain constant for a given category system and for a specified number of sequences to be compared. This means that one factor influencing the outcome of an LRC test is the length of the interaction sequences chosen for the comparison.

Computer Simulation

The Choice of H_1: $(P_{jk1}) \neq (P_{jk2})$. The pair of parentheses indicates that P_{jk1} and P_{jk2} are matrices of transition probabilities. For convenience, the researcher deviated slightly from the subscript notation used in the Section I and shifted to Darwin's.

<u>In terms of educational significance, two interaction sequences, displayed in matrix form as (n_{jk1}) and (n_{jk2}), are said to be educationally homogeneous if the two classroom situations which they represent have the same educational outcomes on the basis of an outside criterion, e.g.: achievement or attitude.</u> Here, homogeneity refers to educational outcome and is to be distinguished from equal transition probabilities. To obtain a pair of unequal sequences for H_1, two identical sequences may be made to vary a little by slightly changing the frequency in one cell (or the frequencies in a few cells) of the matrix. One would judge, without further tests, that the two sets of classroom interaction which the two matrices represent were extremely homogeneous. One could continue modifying cell frequencies and still produce two educationally homogeneous sequences. For the results of this study to be meaningful, one should choose a pair which would satisfy the criterion of being representative of a real situation, one where two teachers are producing the same educational outcomes through similar classroom situations.

The first question this section attempts to answer is: How sensitive is Darwin's test to differences between two classroom interaction sequences? That is, how often does Darwin's test tend to reject the hypothesis that two sequences are equal when it is known that they are homogeneous in terms of educational outcomes?

The answer to these questions can be approximated by choosing a pair of sequences with optimum inequality. That is, the two sequences should represent a pair of realistic classroom situations which are judged to be educationally homogeneous. These two sequences would be made to generate pairs of sequences to be tested for equality by Darwin's LRC, and the percentage of rejection would be observed. This percentage is the proportion of times one would make a "Type I error" <u>in the educational sense</u>, that is, the proportion of times one would reject the hypothesis of educational homogeneity when it is true.

In connection with the problem of choosing H_1, two ideas should be recalled: (1) the underlying objective in this section was to find a way of interpreting any statistical significance in the light of educational significance; (2) the problem in testing differences between two interaction data sets was the high sensitivity of the LRC (statistically significant differences resulted from the test, even though the two sequences being tested were known to be practically the same).

The author created two composite matrices that were based on typical transition probabilities, and in addition, represented homogeneous outcomes. If pairs of interaction sequences were generated from this pair, one would get a distribution of Chi Square values and could see how often the hypothesis of equality would be rejected.

Criteria for selection. The following statements summarize the criteria for choosing the pair of sequences for the alternative hypothesis:

1. The two sequences have no significantly different educational outcomes.

2. The proportions in the cells of the matrices represent some identifiable target population. To achieve these proportions, the researcher used actual data, instead of determining the extent of homogeneity by subjective judgment.

3. The length of each sequence is realistic. Normally, 30 minutes to two hours of classroom observation were made in projects involving interaction analysis.

Source of data for the alternative hypothesis. Flanders [5] observed 16 eighth-grade mathematics classes. The teachers all taught the same two-week unit of study, the materials of instruction being kept constant. Teacher influence was controlled by measuring the spontaneous patterns of teachers, while the adjusted final achievement scores of the students was an outcome variable. By "adjusted final achievement scores" is meant the scores which took into account the initial ability of the students.

Steps in the Computer Simulation. The following is an outline of the steps in generating pairs of sequences and obtaining a probability distribution of Darwin's LRC.

1. The 16 eighth-grade mathematics teachers were first ranked according to the adjusted post-test score. Then the odd-numbered teachers in the list formed one group, and the rest formed the other group. Table 5.1 shows these two groups, with their adjusted post-test scores. The result was a nonsignificant difference between the mean achievement scores of the two groups.

TABLE 4

TWO GROUPS OF 16 EIGHTH GRADE MATHEMATICS TEACHERS, HOMOGENEOUS ON ACHIEVEMENT

Group I		Group II	
Code No.	Achievement Score	Code No.	Achievement Score
V804	34.1	C801	30.7
P805	30.2	D804	29.8
V803	29.1	M802	28.9
A802	27.9	M805	27.3
I801	27.1	T802	26.5
H802	26.3	A803	26.2
E802	24.7	G802	24.1
L803	23.6	G801	21.3
Means	27.9		26.9

2. The matrices of the teachers in each group were then combined to form two composite matrices of the pair--

one composite matrix for each group. The use of several matrices combined, rather than selecting a single matrix of a short sequence, would make the composite matrix more representative of a group of similar classroom situations. In other words, the ML estimates in a composite matrix approximate the averages of the separate ML estimates in the component single matrices.

3. From the two composite matrices, the ML estimate of p_{jkl} was computed. The ML estimate of p_{jkl} is

$$p_{jkl} = \frac{n_{jkl}}{n_{j.l}}, \quad j,k = 1,\ldots,10, \quad l = 1,2. \tag{22}$$

The estimates p_{jk1} and p_{jk2} are given in Table 5. It can be observed that the corresponding transition probabilities of the two matrices (A) and (B) are not very different.

4. Cumulative transition probabilities were computed from the ML estimates. These are shown in Tables 6 and 7.

5. From the cumulative probabilities obtained in

TABLE 5

MAXIMUM LIKELIHOOD ESTIMATES (TIMES 1000) OF TRANSITION PROBABILITIES p_{jk} FOR THE TWO SEQUENCES A AND B OF THE ALTERNATIVE HYPOTHESIS

i \ j		1	2	3	4	5	6	7	8	9	10
1	A	097	001	020	157	373	078	039	039	059	137
	B	077	015	023	101	628	001	078	001	031	047
2	A	002	046	081	167	262	095	002	015	100	229
	B	003	122	143	162	281	027	016	014	097	135
3	A	002	036	304	155	304	030	007	042	073	049
	B	003	023	323	214	249	015	007	027	075	064
4	A	002	004	003	155	051	024	008	635	031	086
	B	001	003	004	141	040	017	004	663	064	064
5	A	001	003	003	083	797	038	013	006	027	028
	B	001	002	003	103	787	025	008	003	039	027
6	A	001	002	001	048	122	323	022	219	057	205
	B	001	001	002	060	122	296	047	170	084	218
7	A	001	003	002	092	206	065	388	033	047	163
	B	001	005	005	066	109	101	388	024	092	208
8	A	001	045	112	139	239	068	031	313	028	024
	B	002	048	214	235	208	065	015	153	027	034
9	A	002	041	165	071	247	085	045	007	278	058
	B	004	032	192	075	224	064	036	002	315	055
10	A	002	011	003	074	113	098	035	033	090	542
	B	001	006	005	070	099	078	058	017	111	555

TABLE 6

CUMULATIVE PROBABILITIES (TIMES 1000)
FOR SEQUENCE A

	1	2	3	4	5	6	7	8	9	10
1	097	098	118	275	647	725	765	804	863	1000
2	002	048	129	296	558	653	655	670	771	1000
3	002	037	341	496	799	829	836	878	951	1000
4	002	006	009	164	215	239	247	882	914	1000
5	001	005	008	091	888	926	939	945	972	1000
6	001	003	004	052	174	497	518	738	795	1000
7	001	004	006	098	304	368	757	790	837	1000
8	001	046	158	297	536	605	635	948	976	1000
9	002	043	209	280	526	612	657	663	942	1000
10	002	012	015	089	202	300	335	368	458	1000

TABLE 7

CUMULATIVE PROBABILITIES (TIMES 1000)
FOR SEQUENCE B

	1	2	3	4	5	6	7	8	9	10
1	077	091	114	215	843	844	921	922	953	1000
2	003	124	268	430	711	738	754	768	865	1000
3	003	026	348	562	811	826	834	861	936	1000
4	001	004	008	150	189	206	210	873	936	1000
5	001	004	007	110	898	922	931	934	973	1000
6	001	002	003	063	185	482	529	698	782	1000
7	001	006	011	077	186	287	675	700	792	1000
8	002	050	263	499	707	772	787	940	966	1000
9	004	036	228	304	528	592	628	630	945	1000
10	001	007	012	081	181	259	317	334	445	1000

step four, 500 pairs of sequences were generated. At the same time, the LRC and its standard deviate z were computed for each generated pair. The procedure for simulation is simply starting a chain with the category on silence, 10. A random number with uniform distribution over the range 0.000 to 1.000 is generated, and the column category in row ten, with a probability greater than or equal to this random number, determines the next category, j. Thus, $n_{10,j}$ is incremented by one, i.e., a tally is made in the (10,j) cell. This new category determines the next row, i (equal to the preceding j). A random number is again obtained to determine the category (j) entry in this row. This cycle continues until the desired length of the sequence is reached. For each pair generated, the LRC and z values were computed. The 500 pairs generated produced 500 LRC and z values.

5. The 500 LRC values were then ranked from smallest to largest in order to locate every fifth percentile for purposes of analysis.

7. Since the LRC is sensitive to the lengths of the

sequences in the pair, data were also generated for different sequence lengths. In this step, only 20 pairs were generated for each length. The lengths tried were 500, 1000, 2000, and 6000 tallies, in addition to the length of 4000 which had been used in generating the 500 pairs.

8. The 20 LRC values in each set were then ranked from the smallest to largest to identify the percentile value for each LRC value.

Results and Discussion

Pairs of Sequences of Length 4000. Table 8 is a cumulative distribution of the 500 values of Darwin's LRC computed from the 500 pairs of sequences generated from the two composite matrices in step two of the preceding subsections. The lowest z value in the generation is 3.622, which is still statistically significant ($p < 0.01$). Hence, if the sample pair being tested is the same type as the original pair--eighth-grade data, homogeneous on achievement, etc.--in most (or all) cases, one would wrongly infer (wrong in the educational sense) that the two sequences in the pair are not equal, i.e., not educationally homogeneous.

It has now been demonstrated that, using two educationally homogeneous sequences, Darwin's test is so powerful that the power is equal to one. For such time as researchers continue to use this test, the investigator suggests reducing the probability of the Type I error

TABLE 8

EMPIRICAL CUMULATIVE DISTRIBUTION[a] OF DARWIN'S LIKELIHOOD RATIO CRITERION AND ITS STANDARD DEVIATE Z

Generation Pair No.	LRC_0	z_0	$P(z > z_0)$
304	164.625	4.755	.95
122	172.445	5.181	.90
23	179.937	5.580	.85
291	183.641	5.775	.80
118	187.742	5.987	.75
14	192.312	6.222	.70
436	195.266	6.372	.65
326	199.102	6.565	.60
461	201.586	6.689	.55
359	204.281	6.823	.50
114	206.391	6.927	.45
378	209.930	7.100	.40
482	212.977	7.249	.35
197	217.008	7.443	.30
130	222.016	7.682	.25
266	225.617	7.852	.20
271	230.859	8.098	.15
221	235.672	8.320	.10
131	247.414	8.855	.05
240	250.445	8.991	.04
223	253.344	9.120	.03
488	257.172	9.289	.02
8	261.375	9.474	.01
104	274.680	10.048	.00

[a]The length of a sequence in the 500 pairs of sequences generated, is 4000 tallies.

in order not to reject a large number of hypotheses which, though statistically false, are educationally true. In other words, the cutoff point in the range of LRC values should be much greater than those values at the usual levels of significance, i.e., greater than the Chi Square values at .05, or .01, or .001. In effect, one reduces the power of the test even as far as .05.

Thus, from Table 8, under the given alternative hypothesis,

$$p(z > 8.855) = .05. \tag{23}$$

That is, if the cutoff point is set at $z = 8.855$, the power is reduced to approximately .05. The value .05 in the table represents the fifth percentile, or the probability of rejecting the hypothesis of equality under the given alternative hypothesis. This probability is the empirical power of the test, at $z = 8.855$, under the alternative hypothesis given by the transition probabilities in Table 5.

In contrast, the table of the normal probability integral---single tail probability [~~10~~ 7]--gives

$$p(z > 4.99) = 0.30190(10^{-6}). \tag{24}$$

The above equation is based on the theoretical distribution of z under the hypothesis of equality of the two sequences. Hence, in the statistical sense, choosing $z = 8.855$ as the boundary of the critical region reduces to near-zero the level of significance α, or the probability of a Type I error.

In summary, the cutoff point of z in rejecting equality can be given by a normal distribution table, but the cutoff point for rejecting educational homogeneity would be suggested by a table such as Table 8.

Figure 1 is a graph of the cumulative empirical distribution of z obtained from Table 8. Under the alternative hypothesis of educational homogeneity given by Table 5, the probability of a z value smaller than 10 is near one. The application of the empirical distribution of LRC values obtained at this stage would be limited, considering the number of assumptions to be met--such assumptions as length of sequence, type of classroom, grade level, subject taught, and educational criterion. However, the values in Table 8 would easily suggest the magnitude of the LRC which one might set as the boundary of the critical region. For purposes of illustration, suppose these assumptions were met by two sequences whose difference was being tested for significance. Suppose, the LRC obtained were 8.33. While a normal probability table indicates this value to be highly significant with α practically zero, Table 8 indicates that rejecting the hypothesis at this point leaves a high probability that the sequences are homogeneous from an educational standpoint.

Pairs of Sequences of Varying Lengths. Table 9 shows the ranked values of LRC and z computed from generated pairs of varying lengths--20 pairs to each given length. For any

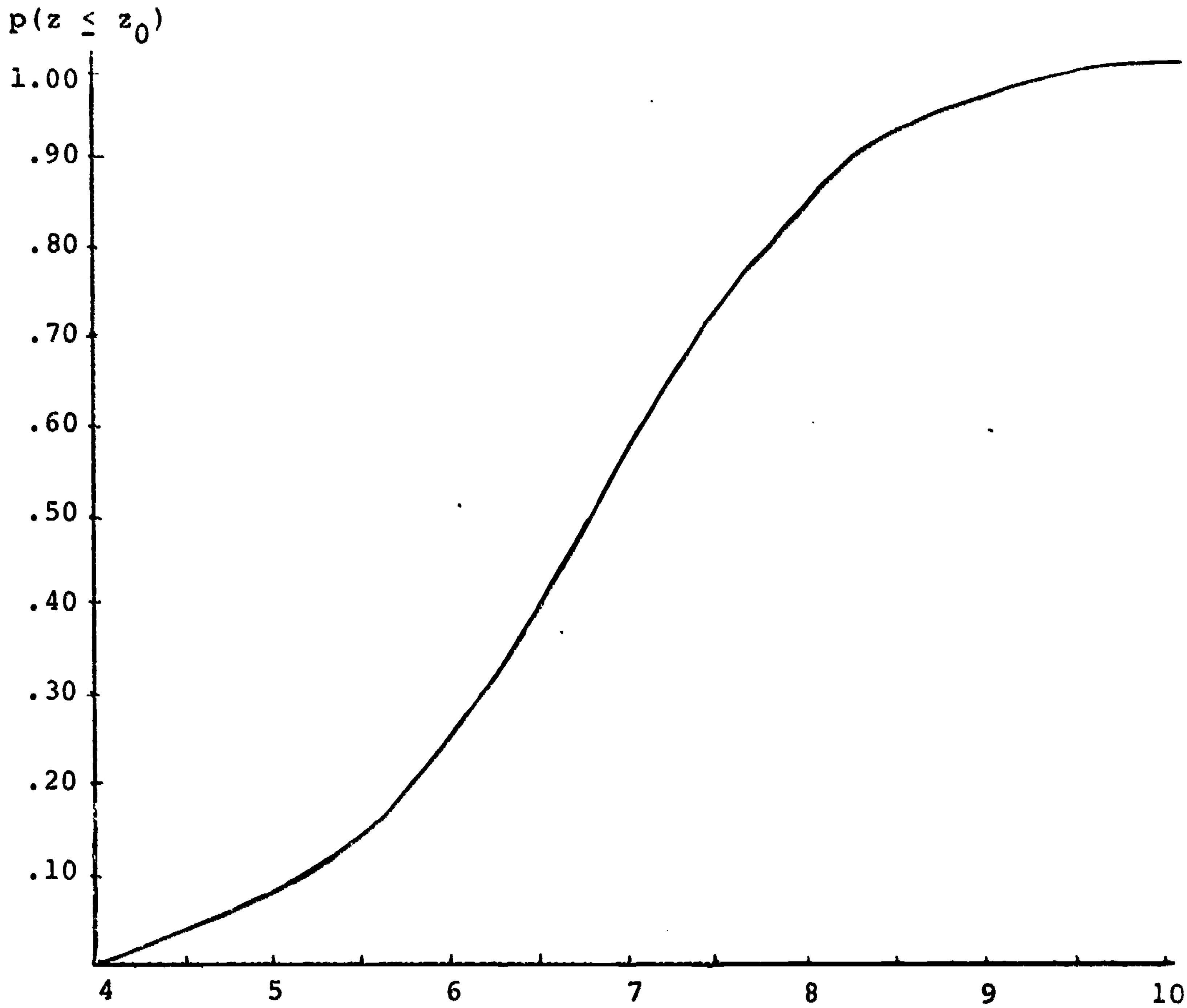

Fig. 1 --Cumulative empirical distribution of Darwin's Likelihood Ratio criterion computed from 500 pairs of sequences, generated from a pair of interaction sequences.*

*Length of each sequence = 4000 tallies.

TABLE 9

RANKED VALUES OF LRC AND Z COMPUTED FROM 20 PAIRS OF SEQUENCES BY LENGTH OF GENERATED SEQUENCE

Rank No.	Sequence Length										Percentile[a]
	500		1000		2000		4000		6000		
	LRC	Z	LRC	Z	LRC	Z	LRC	Z	LRC	Z	
1	68.41	-1.69	87.21	-0.18	96.53	0.50	170.74	5.09	190.25	6.12	.95
2	73.79	-1.24	92.33	0.20	106.33	1.19	177.57	5.46	226.37	7.89	.90
3	74.98	-1.14	92.58	0.22	116.74	1.89	180.86	5.63	229.25	8.02	.85
4	76.63	-1.01	95.88	0.46	119.94	2.10	181.20	5.65	239.12	8.48	.80
5	77.85	-0.91	96.05	0.47	127.59	2.58	181.35	5.65	247.75	8.87	.75
6	78.39	-0.87	96.50	0.50	127.92	2.61	187.48	5.97	250.50	8.99	.70
7	79.40	-0.79	96.89	0.53	128.38	2.63	189.60	6.08	252.62	9.09	.65
8	81.47	-0.63	97.44	0.57	131.62	2.83	189.82	6.09	254.50	9.17	.60
9	83.49	-0.47	100.72	0.80	132.39	2.88	192.31	6.22	254.75	9.18	.55
10	84.34	-0.40	101.41	0.85	133.65	2.96	200.34	6.63	256.12	9.24	.50
11	87.16	-0.19	101.79	0.88	142.62	3.50	202.55	6.74	256.87	9.28	.45
12	87.41	-0.17	102.53	0.93	143.66	3.56	204.74	6.85	260.62	9.44	.40
13	87.41	-0.17	103.30	0.98	147.02	3.76	209.11	7.06	262.50	9.52	.35
14	88.67	-0.07	105.66	1.15	147.33	3.78	217.89	7.49	268.63	9.70	.30
15	89.80	0.01	108.75	1.36	148.43	3.84	223.16	7.74	267.00	9.72	.25
16	91.08	0.11	112.44	1.61	150.35	3.95	223.54	7.75	269.75	9.84	.20
17	98.27	0.63	114.27	1.73	154.29	4.18	231.11	8.11	270.38	9.86	.15
18	100.62	0.80	116.92	1.90	160.55	4.53	236.59	8.36	271.88	9.93	.10
19	100.74	0.80	131.02	2.80	168.75	4.98	254.42	9.17	312.75	11.62	.05
20	102.27	0.91	140.27	3.36	179.48	5.56	261.37	9.47	327.12	12.19	.00

[a]Since 20 pairs were generated for each length of sequence, the 20 ranked values would represent estimates of the percentiles in intervals of five, for each given length.

given percentile rank, the increasing values of LRC for increasing lengths can be noted. Thus, a z of 4.98 would have a significant meaning in education if the number of tallies in each member of the pair is near 2000, but not if the number is near 4000, which requires a z of 9.17 for the difference to have any significance in education.

A Comparison of the Empirical Distribution of Darwin's LRC with the χ^2 Distribution. A χ^2 table gives values at 90 degrees of freedom at certain percentile points. These χ^2 values are posted in the first column of Table 10. The last column gives the percentile points available from a χ^2 table. The remaining columns give the corresponding LRC values obtained from the generated sequences of lengths varying from 500 to 6000. The table shows that the sequence length of 500 has LRC values closest to the tabulated χ^2 values, with discrepancies widening as the critical region (probability of a greater value) decreases. However, since the LRC for length 500 is close to but does not significantly exceed χ^2, the tabulated values of χ^2 could provide a safe guideline for educational inference for sequences of 500 tallies to each member of a pair. This statement is particularly true if the major concern is to minimize the chances of making a Type I error, that is, when one desires a greater margin of safety in being right in rejecting equality.

When these values were plotted on the graph (Figure 2)

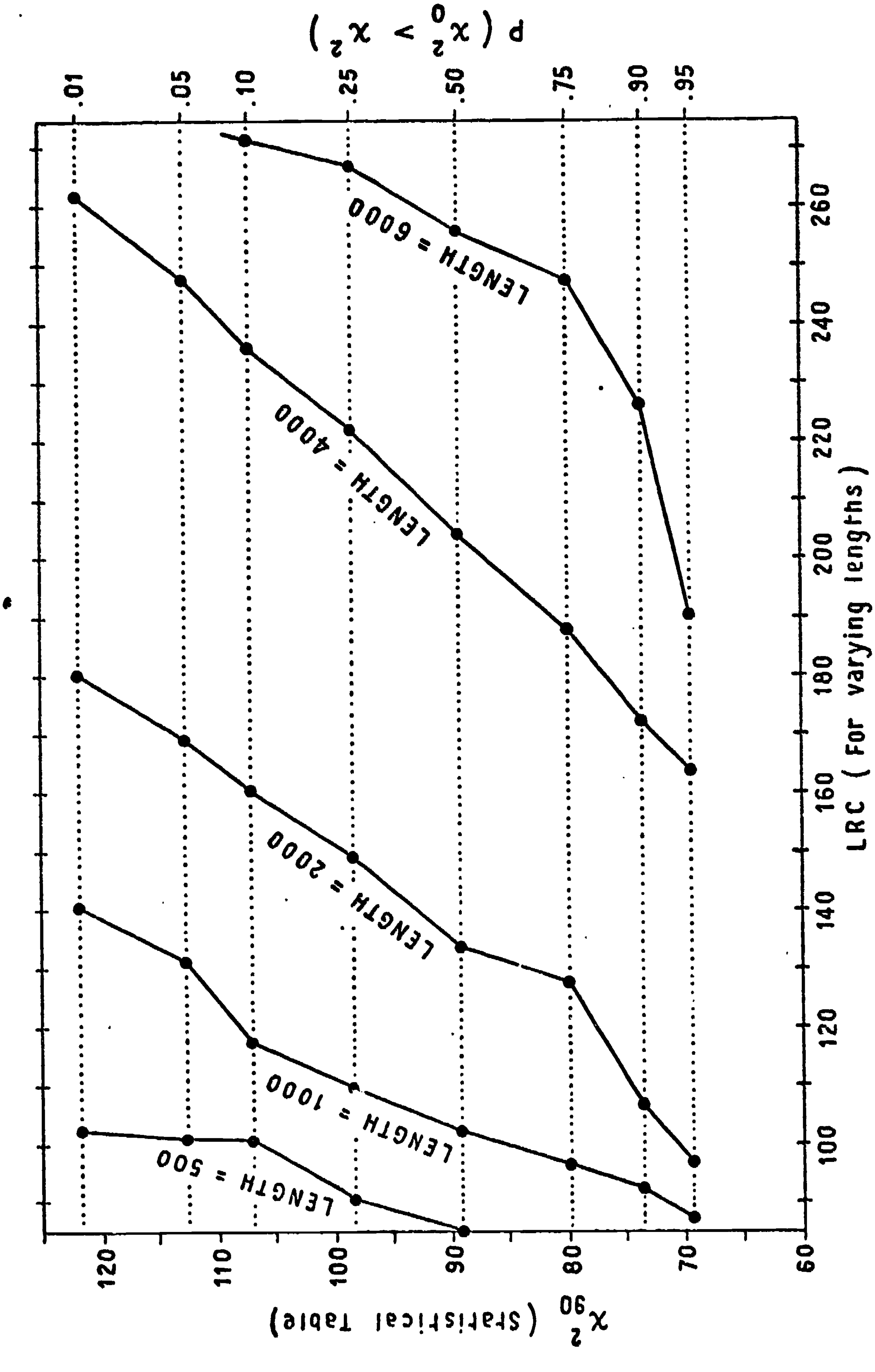

Fig 2 - Graphical relationship between tabulated χ^2 (df = 90) values and LRC obtained from generated pairs of sequences of different lengths at given percentile points

TABLE 10

A COMPARISON OF THE CUMULATIVE DISTRIBUTION OF χ^2_{90} FROM A CHI SQUARE TABLE AND THE EMPIRICAL CUMULATIVE DISTRIBUTION OF DARWIN'S LRC COMPUTED FROM DIFFERENT SEQUENCE LENGTHS

χ^2_{90} (Table)	Darwin's LRC					Probability of a Greater Value
	Length = 500	Length = 1000	Length = 2000	Length = 4000	Length = 6000	
69.13	68.41	87.21	96.53	164.63	190.25	.95
73.29	73.79	92.33	106.33	172.45	226.37	.90
80.62	77.85	96.05	127.59	187.74	247.75	.75
89.33	84.34	101.41	133.65	204.28	256.12	.50
98.64	89.80	108.75	148.43	222.02	267.00	.25
107.56	100.62	116.92	160.55	235.67	271.88	.10
113.14	100.74	131.02	168.75	247.41	312.75	.05
124.12	102.27*	140.27*	179.48*	261.38	327.12*	.01

*Values at .00 percentile in Table 9

under the given percentile points, it became obvious that those obtained from lengths of 4000 had a more stable relationship with the χ^2 distribution. The line of the plotted points is almost a straight line. The values for the extreme lengths of 500 and 6000 are the most erratic, while those for lengths of 2000 and 1000 have only one and two points, respectively (out of eight percentile points), which deviate from a linear trend.

Implication and Possible Extension of this Project

By an empirical procedure, the author obtained the power of Darwin's LRC at one point in the set of all possible alternative hypotheses. Pairs of sequences were generated from only one pair of sequences. If this procedure were to be replicated for other pairs of sequences under different alternative hypotheses and for the same lengths, boundary points corresponding to the different alternative hypotheses would be obtained. Such replication would give more information on the range of the boundary points.

It should be recalled that the transition probabilities on which the simulation was based were calculated from a set of 16 interaction matrices of eighth-grade mathematics classes, split into two subsets of classes having no significant difference in the achievement means. If conditions such as grade level, subject taught, and educational criterion are varied from one alternative hypothesis to another, the simulation would produce approximate LRC and z values which would serve as cutoff points for tests on data sets which have conditions similar to those characterizing the basic data sets in the simulation process.

However, even without the replications suggested here, the distribution of LRC and z values generated from one alternative hypothesis can give the interaction analyst a fair idea of the magnitude of the LRC and z values in comparison with the tabulated theoretical distribution of Chi Square. A user may not necessarily use the values presented in this paper, but, having some empirical values to compare his own results with, he may find in them some basis for his interpretation of the significant results of Darwin's LRC.

TABLE 11

SUGGESTED BOUNDARY POINTS FOR LRC AND z, BASED ON VALUES AT THE 95th AND 99th PERCENTILES OF THEIR EMPIRICAL DISTRIBUTIONS FOR DIFFERENT SEQUENCE LENGTHS

Length of Sequence	95th Percentile		99th Percentile	
	LRC	z	LRC	z
500	100.74	0.80	102.27	0.91
1000	131.02	2.80	140.27	3.36
2000	168.75	4.98	179.48	5.56
4000	247.41	9.17	261.38	9.47
6000	312.75	11.62	327.12	12.19

Thus, the critical values of LRC and z were taken from Table 10 and presented in Table 11 as the suggested boundary points (at the 95th and 99th percentiles of their distributions) for the rejection of the hypothesis. It is left to the reader to use his insight and experience with his data to judge whether these values are relevant to his own research.

LIST OF REFERENCES

1. Flanders, N. A., Project Director, Teacher Influence, Pupil Attitudes and Achievement. Cooperative Research Monograph Series. U.S. Department of Health, Education and Welfare, Office of Education, U.S. Government Printing Office, Washington, D.C. 1965.

2. Feller, William, An Introduction to Probability Theory and Its Applications. John Wiley and Sons, Inc., New York. 1967.

3. Darwin, J. H., Note on the Comparison of Several Realizations of a Markoff Chain. Biometrika, 46:412-419, 1959.

4. Hoel, P. G., A Test of Markoff Chains. Biometrika, 41:430-433. 1954.

5. Flanders, N. A., Basic and Applied Research Proposal, Submitted to the U.S. Commission of Education, for the proposed project: "Teacher Influence Patterns and Pupil Achievement in the Sixth, Fourth, and Second Grade Levels." 1964.

6. Anderson, T. W. and Goodman, Leo A., Statistical Inference about Markov Chains. Ann. Math. Statist. 28(1):89-110. 1957.

7. Fisher, Ronald A. and Yates, Frank, Statistical Tables for Biological, Agricultural and Medical Research, Hafner Publishing Company, Inc. New York. 1963.